

4/5/1 (Item 1 from file: 351)  
DIALOG(R) File 351: Derwent WPI  
(c) 2002 Thomson Derwent. All rts. reserv.

012094254 \*\*Image available\*\*  
WPI Acc No: 1998-511165/ 199844  
XRPX Acc No: N98-398878

**Voiced sound identification method in audio encoder - involves  
identifying sub-frame as voiced sound signal if low frequency band pair  
high frequency band energy ratio of each sub-frame is more than critical  
value**

Patent Assignee: HETAI SEMICONDUCTOR CO LTD (HETA-N)

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 10222194	A	19980821	JP 9732623	A	19970203	199844 B

Priority Applications (No Type Date): JP 9732623 A 19970203

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
JP 10222194	A		8	G10L-009/14	

Abstract (Basic): JP 10222194 A

The method involves dividing an audio frame data of an input audio into four sub-frames. The NC value of each sub-frame is compared with a high critical value and a low critical value respectively. The energy and LSP coefficient value of each sub-frame are distinguished.

The low frequency band pair high frequency band energy ratio (LOH) of each sub-frame is calculated, and it is judged whether the calculated ratio is more than a critical value. If it is more than critical value, then the sub-frame is identified to be a voiced sound signal. If it is not more than critical value, then the sub-frame is identified to be unvoiced sound signal.

ADVANTAGE - Improves speech synthesis quality. Identifies voiced and unvoiced sound correctly.

Dwg.2/3

Title Terms: VOICE; SOUND; IDENTIFY; METHOD; AUDIO; ENCODE; IDENTIFY; SUB; FRAME; VOICE; SOUND; SIGNAL; LOW; FREQUENCY; BAND; PAIR; HIGH; FREQUENCY; BAND; ENERGY; RATIO; SUB; FRAME; MORE; CRITICAL; VALUE

Derwent Class: P86; T01; W04

International Patent Class (Main): G10L-009/14

International Patent Class (Additional): G10L-009/00

File Segment: EPI; EngPI

4/5/2 (Item 1 from file: 347)  
DIALOG(R) File 347: JAPIO  
(c) 2002 JPO & JAPIO. All rts. reserv.

05939094 \*\*Image available\*\*  
DISCRIMINATING METHOD FOR VOICE SOUND AND VOICELESS SOUND IN VOICE CODING

PUB. NO.: 10-222194 A]  
PUBLISHED: August 21, 1998 (19980821)  
INVENTOR(s): RIN SHINTO  
RIN SHINAN

APPLICANT(s): GOTAI HANDOTAI KOFUN YUGENKOSHI [000000] (A Non-Japanese Company or Corporation), TW (Taiwan)

APPL. NO.: 09-032623 [JP 9732623]  
FILED: February 03, 1997 (19970203)

INTL CLASS: [6] G10L-009/14; G10L-009/00

JAPIO CLASS: 42.5 (ELECTRONICS -- Equipment)

JAPIO KEYWORD: R108 (INFORMATION PROCESSING -- Speech Recognition & Synthesis)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-222194

(43) 公開日 平成10年(1998) 8月21日

(51) Int.Cl.<sup>6</sup>

G 1 0 L 9/14  
9/00

識別記号

F I

G 1 0 L 9/14  
9/00

C  
C

審査請求 有 請求項の数 8 F D (全 8 頁)

(21) 出願番号 特願平9-32623

(22) 出願日 平成9年(1997) 2月3日

(71) 出願人 396009779

合泰半導體股▲ふん▼有限公司  
台湾新竹市科学工業園區研新二路5號

(72) 発明者 林 進燈

台湾台南市北區東興里20鄰小東路299巷2號

(72) 発明者 林 信安

台湾宜蘭市吳沙路256號

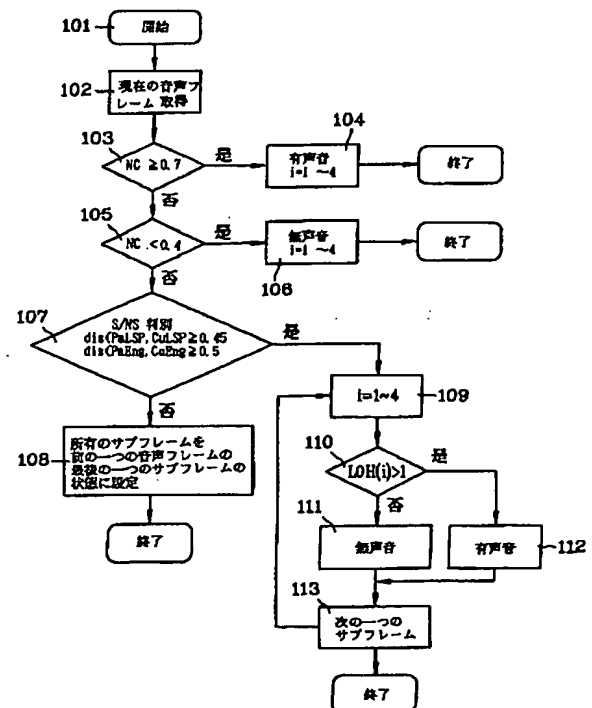
(74) 代理人 弁理士 竹本 松司 (外4名)

(54) 【発明の名称】 音声符号化における有声音と無声音の識別方法

(57) 【要約】

【課題】 音声符号化における有声音と無声音の識別方法の提供。

【解決手段】 入力音声の音声フレームデータを四つのサブフレームに分割し、さらに各サブフレームに対して有声音か無声音かの判別を以下のように行う。各サブフレームのNC値を高臨界値と低臨界値とそれぞれ比較し、安定か不安定かの判別ステップで該サブフレームのエネルギー量値とLSP係数値の大きさをそれぞれ判別し、両者が設定された臨界値より大きければ、サブフレームの低周波数帯域対高周波数帯域エネルギー比率(LOH)の判別ステップを行い、LOH値判別ステップ中で各サブフレームがある臨界値以上か否かを判定し、イエスであればサブフレームを有声音信号と、ノーであれば無声音信号と判別する。



## 【特許請求の範囲】

【請求項1】 一種の音声符号化における有声音と無声音の識別方法であって、入力音声の音声フレームデータの属性を識別するのに用いられる方法であり、該方法は以下のaからfのステップを包括する、

- a. 現在の入力音声の音声フレームデータを四つのサブフレームに分割する、
- b. 四つのサブフレームの正規化相互相関値 (Normalized Cross correlation Value)、即ちNC値が一つの高臨界値以上であるか否かを判別し、もし判別結果がイエスであれば、入力された現在の音声フレーム中の四つのサブフレームがいずれも有声音信号であると判定する、
- c. もし上記bのステップでサブフレームのNC値が高臨界値以上でなければ、該NC値が一つの低臨界値より小さいか否かを判別し、もしイエスであれば、該音声フレーム中の四つのサブフレームがいずれも無声音信号に属すると判定する、
- d. もし上記cのステップでNC値が低臨界値より小さいと判別されたなら、安定か不安定かの判別ステップを執行し、該サブフレームのエネルギー量値と線スペクトル対偶 (line Spectrum Pair) 係数値、即ちLSP係数値の大きさをそれぞれ判別する、
- e. もしエネルギー量値とLSP係数値が設定された臨界値より大きくなければ、音声信号が安定状態を呈すると判定し、四つのサブフレームの属性全部を前の一つの音声フレーム中の最後の一つのサブフレームの有声音か無声音の状態と同じと設定し、
- f. 上述のeのステップ中で、もしエネルギー量値とLSP係数値が設定された臨界値より大きければ、該サブフレームの低周波数帯域対高周波数帯域エネルギー比率 (Low to High Band Energy Ratio Value)、即ちLOH値の判別ステップを執行し、各一つのLOH値に対してある臨界値より大きいかなんかを判定し、もし臨界値より大きければ、該

サブフレームを有声音信号と判定し、もし大きくなければ該サブフレームが無声音の信号と判定し、次の一つのサブフレームに判別を進行して、四つのサブフレームの全てに対する判別を終えて終了する、以上を特徴とする、音声符号化における有声音と無声音の識別方法。

【請求項2】 bのステップでサブフレームのNC値を判別する時に用いられる高臨界値は0.7に設定することを特徴とする、請求項1に記載の音声符号化における有声音と無声音の識別方法。

10 【請求項3】 cのステップでサブフレームのNC値を判別する時に用いられる低臨界値は0.4に設定することを特徴とする、請求項1に記載の音声符号化における有声音と無声音の識別方法。

【請求項4】 dのステップの安定か不安定かの判別ステップ中、サブフレームのエネルギー量値の判別では、前の一つのエネルギー量と現在のエネルギー量の差値が設定されたある臨界値以上であるか否かの判断を行う、請求項1に記載の音声符号化における有声音と無声音の識別方法。

20 【請求項5】 エネルギー量値の判別ステップで、設定された臨界値は0.45とする、請求項4に記載の音声符号化における有声音と無声音の識別方法。

【請求項6】 dのステップの安定か不安定かの判別ステップ中、サブフレームのLSP係数値の判別では、前の一つの平均LSP係数値と現在のLSP係数値との差値を判断することを特徴とする、請求項1に記載の音声符号化における有声音と無声音の識別方法。

30 【請求項7】 サブフレームのLSP係数値の判別ステップで用いる臨界値は0.4と設定することを特徴とする、請求項6に記載の音声符号化における有声音と無声音の識別方法。

【請求項8】 fのステップで、サブフレームのLOH値の判別ステップ中、LOHの定義は以下の数式1とされ、

【数1】

$$LOH(i) = \sqrt{\frac{\frac{1}{W} \sum_{k=-W/2}^{W/2-1} s^2 l_{p} l_k(k + d_{offset}(i))}{\frac{1}{W} \sum_{k=-W/2}^{W/2-1} s^2 h_{p} l_k(k + d_{offset}(i)) + T_{sil}}}$$

その中iは第i個のサブフレームを代表し、 $S_2 l_{p} l_k$ はもとの信号が1k低域フィルタを通過した後に得られる信号を代表し、定義中、音声信号中の1KHzより低いものと1KHzより高いもののエネルギー量比率は、一つのウィンドウ長さWで割られ、そのいわゆるウィンドウ長さWの定義は、  
ピッチがNsubframeより大きい場合は、 $W = \text{pitch}$

ピッチがNsubframe/2以上でNsubframeより小さい場合は、 $W = 2 * \text{pitch}$ であり、その中、Nsubframeはサンプルのサブフレーム長さを示し、LOHの定義中、静音臨界値 $T_{sil}$ は現在の音声フレームの最大音声値であり、該 $T_{sil}$ 値は1KHzの高域フィルタを通過した音声信号のエネルギー量中に加えられ得て、それにより低エネルギー量の有声音号に無声音として選択される傾向を与え、doffset

t(j)は各一つのサブフレームの中心位置で、その定義は、

$$\text{doffset}(j) = \text{Nsubframe} * (j-1/2), \quad j=1 \sim 4$$

その中、jはサブフレームの番号を表示する、以上を特徴とする、請求項1に記載の音声符号化における有声音と無声音の識別方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は一種の音声符号化の技術に関し、特に、音声符号化技術において有声音か無声音かを識別するのに用いられる方法に関する。

【0002】

【従来の技術】音声合成の技術にあつては、線形予測コーディング(Liner Predictive Coding; LPC)の技術が一般に用いられている。このLPCの方法では、LPC-10音声エンコードが、低ビット率の音声圧縮に広く用いられている。一つのLPC音声エンコードについては、いかに正確に入力音声信号が有声音か無声音かを識別するかが重要な課題であった。というのは、この有声音/無声音識別過程が、音声合成の出力品質に大きな影響を与えうるためであった。

【0003】図1に示されるのは、伝統的な音声符号化技術のブロック図である。図中のブロック中には、インパルス列ジェネレータ11(Impulse Train Generator)、ランダムノイズジェネレータ12(Random Noise Generator)、有声音/無声音切り換えスイッチ13(voice/unvoiced Switch)、利得ユニット14(Gain Unit)、LPCフィルタ15(Liner Predictive Coding Filter)、LPCフィルタ制御変数設定ユニット16が含まれる。

【0004】インパルス列ジェネレータ11の発生する周期性インパルス列(Periodic Impuls

$$NC = \frac{\sum_{n=0}^{N-1} s(n)s(n-t)}{\sqrt{\sum_{n=0}^{N-1} s(n)s(n) \sum_{n=0}^{N-1} s(n-t)s(n-t)}}$$

【0007】しかし、不安定な音声信号、即ち臨界値の上下の不確定レベル区域を変動するものについては、そのNC値の臨界値0.4より小さい程度は非常に小さくなり得て、この時、前述の簡易な判別方法では、正確にそれが有声音信号であるか或いは無声音信号であるかを判別できない。ゆえに、実際の応用時には、誤断が発生する恐れがあった。

【0008】上述の問題を克服し、その判別の正確度を増進するために、周知の技術では前述のNC値の判別の外に、さらに音声信号エネルギー量の判別を執行する必要があり、それによって比較的正確な判別をなす目的を

e Train) 或いはランダムノイズジェネレータ12の発生するノイズ信号(White Noise)は、有声音/無声音切り換えスイッチ13による、その入力信号の類型属性によつた適当な選択切り換えを経て、利得ユニット14を経て信号の利得が行われ、以てその信号のレベルが調整される。そしてさらにLPCフィルタ15がLPCフィルタ制御変数設定ユニット16に設定されたLPC変数(LPC Parameters)に基づき、ろ波を執行し、最後に、LPCフィルタ15の出力端より音声出力S(n)を行う。

【0005】前述の音声識別のステップを執行するとき、識別装置は各一つの入力音声の音声フレーム(Speech Frame)に対してその有声音/無声音判別方法、ピッチ周期(Pitch Period)、LPC変数、及び利得値(Gain Value)を更新する。その目的は、入力音声の変化状況に追従できるようにすることにある。現在ある典型的な技術では、各一つの音声フレームは160回のサンプルを包括し、即ち、一つの所定の音声フレームの大きさの中に、0.02秒ごとにサンプルが取られている。

【0006】前述の音声識別において、その中、有声音か無声音かの判別方法については、伝統的には相関ピッチの強度に基づいて判別する方法がとられている。例えば、もし正規化相互相関値(Normalized Cross correlation Value; NC値)が予め設定された臨界値、例えば0.4以上であれば、その音声フレームは正常な音声信号と判定され、このとき、音声合成器が周期パルス列によりLPCフィルタを励起する。その反対に、もしNC値が臨界値0.4より小さい場合、その音声フレームは無声音信号に属すると判別され、音声合成器はランダムノイズジェネレータで該LPCフィルタを励起する。前述のNC値の定義は以下の数式2による。

【数2】

達していた。

【0009】周知の技術にあつては、このほか、もう一種の有声音/無声音識別方法がある。このもう一種の周知の技術に基づき、音声信号エネルギー量を判別する時には、以下の二種の状況を包括する。

a. 音声エネルギー量

一般的には、無声音の音声エネルギー量は有声音のものより低く、そのエネルギーの二乗平均平方根値(RMS)は、以下の数式3で求められる。

【数3】

$$RMS = \sqrt{\frac{\sum_{n=0}^{N-1} s(n)s(n)}{N}}$$

その中、Nは入力音声信号の音声フレーム全体を代表す

$$ZC = \frac{1}{2} \sum_{i=1}^N \text{abs}[\text{sgn}(s(i-1)) - \text{sgn}(s(i))]$$

前述の音声符号化技術中、各一つの音声フレーム中には、160回のサンプルが含まれ、そのビット数の面では、各一つの音声フレームは34ビットのLPC変数と、6ビットのピッチと、1ビットの有声／無声音と、7ビットの利得値の総計48ビットを含む。

【0010】前述のように、音声を符号化するには、入力音声信号が有声音であるか無声音であるかをいかに正確に判別するかが重要な課題となる。そしてその判別過程が音声合成の出力品質に大きく影響する。もし、有声／無声音判別の過程で、無声音が有声音と誤断されたならば、出力された合成音声は唸り声のような音声となり、もし有声音が無声音と誤断されたならば、出力された合成音声は敲撃音のように聞こえる。この問題に対して、前述の伝統的な技術は有効に解決することができなかった。

【0011】さらに、前述の第2種の伝統の技術では、1ビットを以て音声フレーム中の有声或いは無声音の状態を決定しており、以て、有声と無声音間の臨界状態を含蓄せんとしている。このため、音声フレーム全体が臨界区域にあり、有声音か無声音かが判定されていないので、往々にして出力された合成音声は雑音を有するように聞こえた。

【0012】

【発明が解決しようとする課題】上述の周知の技術の欠点から、伝統的な音声符号化技術には改善の必要があることが分かる。このため、本発明の主な目的は、一種の音声符号化の改良技術を提供して、それにより音声符号化の過程で優れた音声合成出力品質を提供することにある。

【0013】本発明のもう一つの目的は、音声符号化中に用いられる、正確に有声音か無声音かの識別方法を提供し、該識別方法により、正確に入力音声信号中の音声フレームが有声音か無声音かを判別できるようにすることにある。

【0014】本発明のさらにもう一つの目的は、一種の四分分割式の有声／無声音判別の方法 (Quarter Voiced/Unvoiced Decision Scheme) を提供することであり、それは、入力音声信号中の各一つの音声フレームを四つのサブフレームに分割し、その後、各一つのサブフレームに対して、

る。

b. ゼロ交差率 (ZC)

その定義は音声フレーム全体の零交差の回数とされ、以下の数式4で求められる。

【数4】

その関連の変数に基づき、総合的に該サブフレームが有声音か無声音かを判定し、その判別の結果により、音声合成出力端にて正確で自然な音声信号出力を行わせる方法とする。

【0015】本発明のさらにもう一つの目的は、一種の、入力音声信号の音声フレーム中の有声／無声音の正確な判別方法を提供することにある。本発明のステップ中、まず、入力音声の音声フレームを四つのサブフレームに分割した後、順に該四つのサブフレームのNC値 (正規化相互相関値) が高臨界値 (例えば0.7) 以上であるか否かを判別する。その後、さらに、該NC値が低臨界値 (例えば0.4) より小さいかを判別する。前述の二つの判別ステップの後、明らかに有声と無声音に属する信号を判別し、続いて、前述の明らかに有声と無声音に属する信号の間に介在する信号を判別する。即ち、もし前述のステップ中、NC値が低臨界値より小さいと判別されたならば、安定／不安定の判別ステップを執行し、該サブフレームのエネルギー量値と線スペクトル対偶 (Line Spectrum Pair; LSP) 係数値、即ちLSP係数値の大きさをそれぞれ判別し、もしエネルギー量値とLSP係数値が予め設定された臨界値より大きくなければ、音声信号が安定状態を呈すると判定し、四つのサブフレームの属性全てを前の一つの音声フレーム中の最後の一つのサブフレームの有声／無声音状態と同じと設定し、もし前述のステップ中、エネルギー量値とLSP係数値が設定された臨界値より大きいと判別されたなら、該サブフレームの低周波数帯域対高周波数帯域エネルギー比率 (Low to High Band Energy Ratio Value; LOH) の判別ステップを執行し、各一つのサブフレームのLOH値が一つの臨界値より大きいかな否かを判定し、もし臨界値より大きければ、該サブフレームを有声音信号と判定し、もしそうでなければ該サブフレームを無声音信号と判定する。同様に次の一つのサブフレームに対して判別を進行し、こうして四つのサブフレーム全てに対する判別を行う。

【0016】

【課題を解決するための手段】請求項1の発明は、一種の音声符号化における有声音と無声音の識別方法であって、入力音声の音声フレームデータの属性を識別するの

20

30

40

50

に用いられる方法であり、該方法は以下のaからfのステップを包括する、

- a. 現在の入力音声の音声フレームデータを四つのサブフレームに分割する、
  - b. 四つのサブフレームの正規化相互相関値 (Normalized Cross correlation Value)、即ちNC値が一つの高臨界値以上であるか否かを判別し、もし判別結果がイエスであれば、入力された現在の音声フレーム中の四つのサブフレームがいずれも有声音信号であると判定する、
  - c. もし上記bのステップでサブフレームのNC値が高臨界値以上でなければ、該NC値が一つの低臨界値より小さいか否かを判別し、もしイエスであれば、該音声フレーム中の四つのサブフレームがいずれも無声音信号に属すると判定する、
  - d. もし上記cのステップでNC値が低臨界値より小さくないと判別されたなら、安定か不安定かの判別ステップを執行し、該サブフレームのエネルギー量値と線スペクトル対偶 (line Spectrum Pair) 係数値、即ちLSP係数値の大きさをそれぞれ判別する、
  - e. もしエネルギー量値とLSP係数値が設定された臨界値より大きくなければ、音声信号が安定状態を呈すると判定し、四つのサブフレームの属性全部を前の一つの音声フレーム中の最後の一つのサブフレームの有声音か無声音の状態と同じと設定し、
  - f. 上述のeのステップ中で、もしエネルギー量値とLSP係数値が設定された臨界値より大きければ、該サブフレームの低周波数帯域対高周波数帯域エネルギー比率 (Low to High Band Energy Ratio Value)、即ちLOH値の判別ステップを執行し、各一つのLOH値に対してある臨界値より大きいと否かを判定し、もし臨界値より大きければ、該サブフレームを有声音信号と判定し、もし小さくなければ該サブフレームを無声音の信号と判定し、次の一つのサブフレームに判別を進行して、四つのサブフレームの全てに対する判別を終えて終了する、
- 以上を特徴とする、音声符号化における有声音と無声音の識別方法としている。

【0017】請求項2の発明は、bのステップでサブフレームのNC値を判別する時に用いられる高臨界値は0.7に設定することを特徴とする、請求項1に記載の音声符号化における有声音と無声音の識別方法としている。

【0018】請求項3の発明は、cのステップでサブフレームのNC値を判別する時に用いられる低臨界値は0.4に設定することを特徴とする、請求項1に記載の音声符号化における有声音と無声音の識別方法としている。

【0019】請求項4の発明は、dのステップの安定か

不安定かの判別ステップ中、サブフレームのエネルギー量値の判別では、前の一つのエネルギー量と現在のエネルギー量の差値が設定されたある臨界値以上であるか否かの判断を行う、請求項1に記載の音声符号化における有声音と無声音の識別方法としている。

【0020】請求項5の発明は、エネルギー量値の判別ステップで、設定された臨界値は0.45とする、請求項4に記載の音声符号化における有声音と無声音の識別方法としている。

10 【0021】請求項6の発明は、dのステップの安定か不安定かの判別ステップ中、サブフレームのLSP係数値の判別では、前の一つの平均LSP係数値と現在のLSP係数値との差値を判断することを特徴とする、請求項1に記載の音声符号化における有声音と無声音の識別方法としている。

【0022】請求項7の発明は、サブフレームのLSP係数値の判別ステップで用いる臨界値は0.4と設定することを特徴とする、請求項6に記載の音声符号化における有声音と無声音の識別方法としている。

20 【0023】請求項8の発明は、fのステップで、サブフレームのLOH値の判別ステップ中、LOHの定義は以下の数式1とされ、

【数1】その中iは第i個のサブフレームを代表し、S<sub>2lplk</sub>はもとの信号がlk低域フィルタを通過した後に得られる信号を代表し、定義中、音声信号中の1KHzより低いものと1KHzより高いもののエネルギー量比率は、一つのウィンドウ長さWで割られ、そのいわゆるウィンドウ長さWの定義は、ピッチがN<sub>subframe</sub>より大きい場合は、W=ピッチ (pitch)

30 ピッチがN<sub>subframe</sub>/2以上でN<sub>subframe</sub>より小さい場合は、W=2\*ピッチ であり、その中、N<sub>subframe</sub>はサンプルのサブフレーム長さを示し、LOHの定義中、静音臨界値T<sub>sil</sub>は現在の音声フレームの最大音声値であり、該T<sub>sil</sub>値は1KHzの高域フィルタを通過した音声信号のエネルギー量値に加えられる得て、それにより低エネルギー量の有声音号に無声音として選択される傾向を与え、doffset(j)は各一つのサブフレームの中心位置で、その定義は、

40 
$$\text{doffset}(j) = N_{\text{subframe}} * (j-1 / 2), \quad j = 1 \sim 4$$

その中、jはサブフレームの番号を表示する、以上を特徴とする、請求項1に記載の音声符号化における有声音と無声音の識別方法としている。

【0024】

【発明の実施の形態】本発明の判別方法では、入力音声信号の音声フレームを4個のサブフレーム (Subframe) に分割し、その後、各一つのサブフレームに対して相関する変数に基づき、相対的に各一つのサブフレームが有声音か無声音かの判別を行う。前述の変数は、

NC、エネルギー量、線スペクトル対偶係数 (line Spectrum Pair; LSP)、及び低周波数帯域対高周波数帯域エネルギー比率 (Low to High Band Energy Ratio Value; LOH) を含む。

【0025】以下は本発明の判別ステップである。図2に示されるのは、本発明の判別フローチャートである。そのステップは以下を包括する。フローチャートの開始ステップ101の後、ステップ102を執行する。ステップ102では現在の音声フレームデータを取得する。続いて、NC値が一つの高臨界値0.7以上であるか否かの判別のステップ103を執行する。該NC値の定義については前述の説明を参照されたい。もし判別結果がイエスであるならば、ステップ104を執行する。ステップ104ではこの入力された現在の音声フレームデータ中の四つのサブフレームがいずれも有声音信号であるか否かを判定し、その後、判別プロセスを終了する。

【0026】もし前述のステップ102中で、NC値が高臨界値0.7以上でないと判別したならば、続いてステップ105で該NC値が低臨界値0.4より低いかなかを判別し、もしイエスであれば、該音声フレーム中の四つのサブフレームがいずれも無声音信号に属すると判定し、その後、判別プロセスを終了する。

【0027】前述のステップ102、103の判別の後、明らかに有声音ないし無声音に属する信号が判別される。続いて、前述の明らかに有声音ないし無声音に属する信号の間に介在する信号を判別し、この一つの不安定な、遷移領域の中、単独のステップ102、ステップ103中のNC値判断ステップにより有声音／無声音の正確な判別を行うことは不可能であり、このため以下の判別方法により本発明の課題を解決することができる。ゆえに以下の判別ステップが本発明の特徴を極めて示すキーステップであるといえる。

【0028】もし前述のステップ105中でNC値が0.4より小さくないと判定されたなら、安定か不安定かの判別ステップ (Stationary/nonstationary Decision; S/NS Decision) を執行する。この一つのステップ中には、二つの判別項目が含まれ、その中の一つは、エネルギー量の判別であり、それは、一つのエネルギー量 (Previous Energy) と現在のエネルギー量 (Current Energy) の差値、即ちdis (PrEng, CuEng) の判別である。さらにもう一步S/NSの判別の正確度を増すために、この一つのステップ中にはさらにLSP係数の判別が包括される。このLSP係数はLPC等化器より取得する。このLSP係数の判別では、前の一つの平均LSP (Past average LSP) と現在LSP (Current LSP) の差値、即ちdis (PaLSP, CuLSP) を取得する。ステップ107のS/NS判別ステ

ップ中、

a. dis (PrEng, CuEng) が0.45以上であり、且つ、

b. dis (PaLSP, CuLSP) が0.4以上であるか否かを判定し、

もし結果がノーであれば、音声信号は安定状態にあることを示し、ステップ108を執行し、四つのサブフレームの属性が全て前の一つの音声フレーム中の最後の一つのサブフレームの有声音又は無声音状態と同じと設定する。反対に、もしステップ107の差値の判別ステップ中で、結果がイエス (即ちエネルギー量或いはLSP係数の変化が極めて速いことを示す) の場合、LOHの判別ステップ (ステップ109から113) を執行し、各一つのサブフレームに対して有声音か無声音かの分類判別を行い、以て正確な判別結果を得る。いわゆるLOHの判別の定義は以下の数式1のとおりである。

【数1】その中iは第i個のサブフレームを代表し、S<sub>2plk</sub>はもとの信号が1k低域フィルタを通過した後に得られる信号を代表する。定義中、音声信号中の1KHzより低いものと1KHzより高いもののエネルギー量比率は、一つのウインドウ長さWで相互に相除され、そのいわゆるウインドウ長さWの定義は以下のとおりである。

W=ピッチ (pitch)      ピッチがNsubframeより大きい場合

W=2\*ピッチ      ピッチがNsubframe/2以上でNsubframeより小さい場合

その中、Nsubframeはサンプルのサブフレーム長さを示す。このほか、LOHの定義中、一つの静音臨界値Tsilを現在の音声フレームの最大音声値として選択し、該Tsil値は1KHzの高域フィルタを通過した音声信号のエネルギー量中に加えられ得て、それにより低エネルギー量の有声音信号が無声音として選択される傾向が得られる。doffset(j)は各一つのサブフレームの中心位置で、その定義は、

$$\text{doffset}(j) = \text{Nsubframe} * (j - 1 / 2), \quad j = 1 \sim 4$$

その中、jはサブフレームの番号を表示する。本発明のLOH判別フローチャート中、ステップ110では先に第1個のサブフレームのLOH (前述の定義を参照) が1より大きいかなかを判別し、もしイエス (1より大きい) ならば、ステップ112を執行し、該サブフレームが有声音信号であると判定する。もしノーであれば、ステップ111を執行し、即ち該サブフレームが無声音の信号であると判定する。その後、さらにステップ113及び119に戻り、次の一つのサブフレームに対して判別を進行し、四つのサブフレーム全てに対する判別を終えて終了する。即ち、上述のLOH判別の後、各一つのサブフレームのLOH値が、もし一つの臨界値より大きければ、該サブフレームは有声音と判定され、大きくなけ

11

れば無声音と判定される。一つの音声フレームの四つのサブフレーム全部の判定が終了した後、結果に基づき符号化する過程に進む。本発明中、四つのサブフレームがただ3ビットを以て符号化され、それは図3に示されるとおりである。その中、1は有声音を示し、0は無声音を示す。

【0029】図3に示される索引値を得た後、対応する値を保存し、符号化の過程を完成し、その後、実際の応用では、周知の音声合成技術を以て復号し、必要な合成音声が発生する。

【0030】

【発明の効果】本発明は、一種の音声符号化の改良技術を提供しており、それにより音声符号化の過程で優れた音声合成出力品質が得られる。本発明はさらに、音声符号化中に用いられる、正確に有声音か無声音かの識別方法を提供しており、該識別方法により、正確に入力音声信号中の音声フレームが有声音か無声音かを判別できる

12

ようになる。本発明はさらにまた、一種の四分割式の有聲／無聲音判別の方法 (Quarter Voiced / Unvoiced Decision Scheme) を提供している。

【図面の簡単な説明】

【図1】伝統的な音声符号化技術の基本ブロック図である。

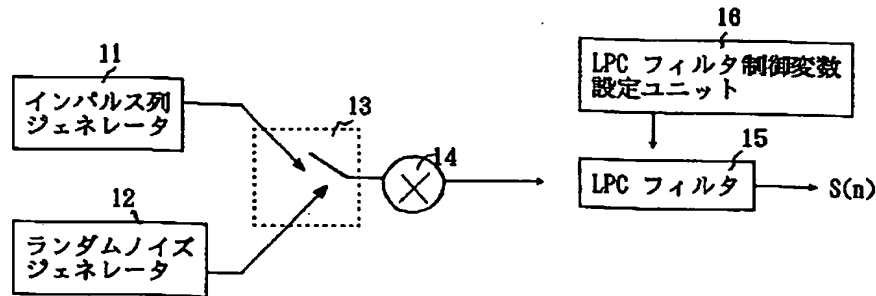
【図2】本発明の判別フローチャートである。

【図3】本発明中、四つのサブフレームを3ビットを以て符号化したコード表である。

【符号の説明】

- 11 インパルス列ジェネレータ
- 12 ランダムノイズジェネレータ
- 13 有聲／無聲音切り換えスイッチ
- 14 利得ユニット
- 15 LPCフィルタ
- 16 LPCフィルタ制御変数設定ユニット

【図1】



【図3】

四つのサブフレームの有聲／無聲音の状態	有聲／無聲音索引値
0000	0
0001	1
0011	2
0111	3
1111	4
1110	5
1100	6
1000	7



【図2】

